DOCUMENT RESUME

ED 413 343 TM 027 635

AUTHOR	Ryan, Katherine E.; Chiu, Shuwan
TITLE	An Examination of Item Context Effects, DIF, and Gender DIF.
PUB DATE	1997-03-00
NOTE	34p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	College Freshmen; *Context Effect; Higher Education; *Item Bias; Item Response Theory; Mathematics Tests; *Sex Differences; Test Format; Test Items
IDENTIFIERS	Item Bias Detection; *SIBTEST (Computer Program)

ABSTRACT

The relationship between differential item functioning (DIF)and test item context effects was studied in an investigation of whether the patterns of gender DIF in parcels of items are influenced by changes in item position, as seen in two forms of a test. A second aim was to determine whether performance of male and female test takers is differentially affected by variation in item position. Items were studied collectively to detect differential bundle functioning (DBF) within content areas of the Midwestern Mathematics Placement Exam, a test for college freshmen. Data from 5 test administrations and 2 test forms for over 5,000 students were used. The simultaneous item bias statistic and the bias estimator, beta, were calculated with the SIBTEST computer program. There were changes in the amount of gender DIF present when the SIBTEST results for forms 1 and 2 were compared. Results from the DBF analyses suggest that analytic geometry items were differentially easier for women on form 2. The idea of DBF is a useful addition to the study of DIF in that this approach provides added power for detecting patterns of differences. An appendix describes test item attributes.(Contains 5 tables and 27 references.) (SLD)

AN EXAMINATION OF ITEM CONTEXT EFFECTS, DIF, AND GENDER DIF

Katherine E. Ryan
Shuwan Chiu
Division of Measurement and Evaluation
University of Illinois
247 Armory Building
505 East Armory
Urbana, IL 61801

For presentation at the Annual Meeting of the American Educational Research Association, Chicago, IL, March, 1997

AN EXAMINATION OF ITEM CONTEXT EFFECTS, DIF AND GENDER DIF

INTRODUCTION

Item Context Effects

Research on item context effects (whether the performance of test items change when the content, difficulty, or order of previous items is altered) has a long and diverse history in educational measurement. The fundamental question in this line of research is whether the cognitive tasks presented on a norm-referenced test are the same for all testtakers (Leary & Dorans, 1985). The work in this area includes investigations of several issues: item order effects (e.g., random rearrangement, section rearrangement), altering item context (changing the content or difficulty of previous items), item order by anxiety, gender, and other effects, section placement, item parameter invariance, and item arrangement effect on score equating (See Hambleton & Traub, 1974; Leary & Dorans, 1985; Plake, Ansorge, Parker, & Lowry, 1982; Wise, Chia, & Park, 1989; Yen, 1980; Zwick, 1991). In an extensive review of the topic by Leary & Dorans (1985), they conclude while there is some evidence of item context effects, the importance of these effects is not well understood.

Since that review was published, the importance of item context effects on item parameter invariance and score equating is clear. These effects must be accounted for in the measurement process (Wise et al., 1989; Yen, 1980; Zwick, 1991). However, the consequences of item context effects for measurement theory and practice in other areas is less apparent.

Context Effects and Gender

One concern is whether or not specific groups of examinees are differentially affected by changes in item position (Leary & Dorans, 1985). Findings from two studies suggest that the impact of altering the item context impacts low achieving testtakers more than high achieving test takers (Wise et al., 1989). Investigators have also explored whether there was an interaction between gender and item order effects on mathematics tests (Hambleton & Traub, 1974; Plake et al., 1982). Results of these investigations were mixed. Plake et al., (1982) examined the effects of three item arrangements on a test of mathematics: easy-to-hard, spiral cyclical (four five item cycles of increasing difficult items), and random. They found an effect; females scored lower than males when the items were ordered easy to difficult with a slightly speeded test.

Plake and her associates also investigated the relationship between differential item performance (items on which males and females perform differently) (DIP) and the effects of item arrangement for males and females in a later study (Plake, Patience, & Whitney, 1988). Using 20 items from the Tests of General Education Development (GED) mathematics item trials, three item arrangements were studied: easy to hard (N=256); easy to hard within content area (N=262) and spiral cyclical (N=261). Only a few significant differences between male and female test takers were found using a modified one parameter IRT approach for detecting DIP.

DIF and Gender DIF

The origins of examining performance differences in achievement are based on what were considered to be bias issues. However, there has

been a considerable change shift in what "biased" means over the last decade. Because of the connotations of the term "bias," in the mid 80s, the more neutral term "differential item functioning" (DIF) was proposed (Holland & Thayer, 1986). In contrast to DIP, DIF refers to items that do not function the same for comparable members of different groups. More recently, the concept of differential bundle functioning (DBF) (a collection of DIF items with a common dimension such as content that collectively produce a bundle of items that are differentially easier for one matched group of test takers in comparison to another) was presented (Stout & Roussos, 1995). Methods that detect only DIF may miss an important phenomena: DIF amplification (Nandakumur, 1993). DIF amplification is the study of a set of DIF items collectively that favors one group in comparison to another at the test score level; these DIF items may show minimal or no DIF when tested as individual items.

Gender DIF and Mathematics Items

Previous work has suggested several factors that influence gender differences in performance on quantitative or mathematics items (Burton, 1996; Lane, Wang, & Magone, 1996; O'Neil & McPeek, 1993; Ryan & Fan, 1996). Doolittle and Cleary (1987) found differences in item functioning were related to item content (e.g., algebra items were differentially easier for females) and item type (word problems were differentially more difficult for females) on the ACT Assessment Mathematics Usage Test (ACTM).

Harris and Carlton (1993) investigated several factors, including item content and format. In their investigation of gender DIF on the SAT, Harris and Carlton (1993) found that after controlling on total test score with the Mantel-Haenzel (MH) procedure, there were systematic

patterns of differences in how males and females performed on the overall test. Applied items (word problems) were differentially easier for male testtakers. Females found geometry and arithmetic items, as well as items requiring higher level thinking skills (non-routine problems or items requiring higher mental processes to be solved) to be differentially more difficult. They also investigated items involving three different categories of items involving visual/spatial factors. Males found items containing figures, graphs, or tables differentially more difficult. There were no significant differences in how items functioned for male and female testtakers on item categories involving a spatial component or figure.

This Study

No studies have examined the relationship between differential item functioning (DIF) and item context effects. The purpose of this study is twofold. First, this investigation examined whether the patterns of gender DIF present in parcels of items is influenced by changes in item position. Second, whether female and male testakers' performance, respectively is differentially affected by variation in item position will be investigated.

Items were studied collectively to detect differential bundle functioning (DBF) within content areas on a test of mathematics for college freshman. Drawing on Doolittle and Cleary's work and Harris and Carlton's work, the relationship between DBF and specific item characteristics identified in previous work, (e.g., such as word problems, items requiring higher-order thinking skills, items containing figures or graphs) was investigated. However, in contrast to previous work in this area, the item categories used in this investigation were

based on the Rule-Space Model developed by K. Tatsuoka (1993). This model was adapted for reporting the math proficiencies for the new Scholastic Assessment Test (SAT-M) (Harnisch, Tatsuoka, & Wilkins, 1995). Items are inspected in relationship to a set of attributes which are the cognitive skills necessary to answer the test question correctly. (See Appendix A for a list of attributes called math challenges and an adaptation used in this investigation.).

The simultaneous item bias test (SIBTEST) was used to study DBF (not bias). SIBTEST, which was proposed by Shealy and Stout (1993) is formulated within a multidimensional IRT perspective and is designed to detect both DIF and DBF.

METHOD

Test

Data from the Midwestern Mathematics Placement Exam (MMPE) exam were used in this investigations. This test is based on course content covered in pre-calculus college courses. While all in-coming freshman with three years of high school mathematics are be administered the test, the purpose of the test is to place students in a pre-calculus course and a first semester calculus course. The test is a 'low stakes assessment.' Students are not required to follow course placement recommendations based on the MMPE test score results. Nevertheless, accurate course placement is useful and efficient for students, faculty, and the institution (Ryan & Fan, 1993). Fairness is also a concern; particularly in light of recent research which suggests that female performance in college mathematics courses is under predicted by college entrance exams entrance examinations like the Scholastic Aptitude Test-Mathematics (SAT-M) (Bridgman & Lewis, 1996; Wainer & Steinberg, 1993).

The test is composed of algebra (18), trigonometry (12), geometry (5) and analytic geometry (5) items. To investigate order effects, two forms of the test were assembled from the item pilot statistics: Form 1, Easy to Difficult; and Form 2, Easy to Difficult within Content Area. The forty items were arranged from easiest to most difficult without regard to content for Form 1. The items were assembled according to difficulty within content area for Form 2. The content areas were presented in the order in which it is typically taught in high school: algebra, geometry, trigonometry, and analytic geometry.

Design and Sample:

Data used in this investigation were collected from five Placement and Proficiency operational tests administrations of incoming-freshman in Spring, 1996. Form 1 was administered during the first three testings; for the last two administrations, the Form 1 and Form 2 test booklets were spiraled to create equivalent groups for data collection. Over five thousand testtakers participated with 3932 examinees completing Form 1 to 1074 testtakers taking Form 2. Test instructions indicated students were allowed 75 minutes to complete the test and that there was no penalty for incorrect answers.

Estimates for coefficient alpha were approximately .88 for both forms. A summary of the descriptive statistics for the total sample and by gender for is reported in Table 1. There is more than a .6 SD difference between male and female test takers in math performance. There are minimal differences in performance between Form 1 and Form 2 testtakers.

Insert Table 1

Item Categorization

The items were categorized by the second author of this paper. She received training on how to analyze items from one of the authors (D. Harnisch) of the attribute analysis for reporting math proficiencies (Harnisch et al., 1995). She has analyzed several sets of mathematics items. After revising the list of cognitive attributes (1-19) in Appendix A, the items were categorized. Seven item categories were constructed from the list of attributes: word problems (1, 5), figures/graphs present (4, 11), construction of graphs (4, 12) algebra operations (1, 3, 9); higher order thinking skill within content area algebra (16), trigonometry (10, 16), and analytic geometry (3, 4, 8, 10). For example, the word problems are based on attributes 1, 5, and 6. Results on previous research had shown that these item types tended to display gender DIF (Doolittle & Cleary, 1987; Harris & Carlton, 1993). The geometry item category was formed from the test specifications developed by the Mathematics Test Development Committee.

The studied item sets were not necessarily independent. There was overlap in the item categorization. For example, two analytic geometry items (38, and 39; Form 2) were studied in two different item categories: Higher order thinking analytic geometry items and Construction of Figures/graphs (items which required the testtakers to construct a graph or figure).

Dimensionality Analyses

Because SIBTEST is conceptualized within an IRT multidimensional perspective, the dimensionality of the studied items were also inspected. DIMTEST was used to examine the dimensionality of the responses for both forms of the test (Nandakumur, 1991; Stout, 1987).

This is a procedure designed to assess whether a set of items meets the requirements for essential dimensionality (essentially one dominant dimension underlying item responses). Stout's statistic T is used to test the null hypothesis of essential dimensionality. To apply DIMTEST, the item pool is split into three subtests. Assessment test 1 (AT1) is used for computing Stout's T. The purpose of assessment test 2 (AT2) is to correct statistical bias due to short test length and/or difficulty differences in AT1. The partitioning test (PT) is used to categorize test takers into subgroups (Nandakumur, 1991).

The dimensionality of the responses from each test form were studied individually and then combined.[1] A backwards procedure was used for testing. The results of the DIMTEST analyses are presented in Table 2. First, because each of the content subtest items could be dimensionally distinct, the items from each content area were inspected, one content area at a time. For example, for the first step, the geometry items were specified as AT1 and the algebra items as AT2 and PT. Then the dimensionality of the geometry items was tested; the null hypothesis was not rejected suggesting the number of dominant dimensions is equal to 1 ($T=-.27$; $p=.61$). Second, the geometry items and algebra items were then pooled and specified as AT2 and PT; the analytic geometry items were specified as AT1. The null hypothesis was not rejected ($T=-.73$; $p=.77$). Third, the trigonometry items were specified as AT1 and the rest of the items were designated as AT2 and PT; the null hypothesis was rejected ($T=5.07$; $p < .05$).

[1] The results of the DIMTEST analyses for Form 2 were not statistically significant, indicating the responses were essentially unidimensional (not reported). The findings for Form 1 parallel the results found when the forms are combined (Form 1 and Form 2 responses) which are described in this section.

Insert Table 2

To further explore the dimensionality of the responses to the trigonometry items, 'agglomerative hierarchical cluster analysis' (HCA), was used (Roussos, 1992). HCA was recently recommended by Douglas, Roussos, and Stout (1996) as an effective approach for identifying dimensionally distinct sets of items when used in combination with DIMTEST. Based on the results of the HCA, the trigonometry items were broken into three item clusters (not reported). Each cluster was specified individually as AT1; all other items were specified as AT2 and PT. The first cluster was tested; the null hypothesis was not rejected (not reported). Those items were then added to AT2 and PT. The procedure was repeated with the second and third subsets of trigonometry items. Only the results for the second subset were statistically significant ($T= 1.88$; $p<.05$). Consequently, these items (7, 24, 25, 38, Form 1; items 25, 27, 28, 29 Form 2) were deleted from further analyses.

SIBTEST

The Simultaneous Item Bias (SIB) statistic and the bias estimator, beta (Shealy & Stout, 1993) were calculated with the SIBTEST computer program (Stout & Roussos, 1995). (See Shealy and Stout (1993) for a discussion of the theory, derivations, and calculations of SIB and beta). Under the null hypothesis of no differential item functioning, a two-tailed hypothesis test, SIB-p ($z= 1.96$, $p < .05$) is conducted to detect uniform DIF for either the focal or reference group. If there is an apriori hypothesis about the direction of the differential functioning, a one-tailed test can also be conducted. SIB and beta can be calculated for each item or a specified set of items. Beta is

interpreted as the difference in the expected total score between the focal and reference groups. For example, if beta is .08, the reference group members have an expected proportion correct for the items (DBF) that is .08 greater than that of comparable focal group members. In the case of DIF, if beta is .08, the reference group has a probability of getting the item correct that is .08 greater than that of matched focal group members. A beta value of .10 is approximately equal to a MH delta difference (MH D-DIF) of -1.0. (See Nandakumur, 1993).

Differential Bundle Functioning Analyses

Table 3 provides a description of the item sets studied in this investigation. Within each item category four analyses were conduct (A, B, C, and D). The total score on all items not under study (excluding the trigonometry items that were dropped) was the matching criterion. For example, for the SIBTEST analysis conducted in Set A for test form 1, the word problems items were designated as the studied item set to test for DIF amplification. The matching criterion was total score on the items not under study: the items that were not word problems.

For the sets of analyses designated as A and B, males were designated as the reference group; females test takers were denoted as the focal group members. The purpose of the analyses in Sets A and B is to examine whether the amount of gender DIF present in parcels of items is influenced by changes in item position. Sets A and B are confirmatory DBF analyses based on item content areas and item categories identified in previous research as a source of gender DIF. Consequently, one-tailed tests, based on apriori hypotheses were conducted.

Whether female or male testakers' performance is differentially affected by variation in item position is investigated in the Sets C and

D analyses. The males who answered Form 1 test items served as reference group members for the analyses from Set C. The Form 2 male examinees were selected as the focal group. The design of the Set D analyses were parallel to Set C, except that female Form 1 and Form 2 test takers were designated as the reference and focal group members, respectively. While the hypotheses for these analyses are also based on earlier research investigating gender DIF (Doolittle & Cleary, 1987; Harris & Carlton, 1993), these are primarily exploratory DBF analyses. Therefore, two tailed hypotheses tests were conducted.

Insert Table 3 here

RESULTS

Confirmatory DBF Analyses

The results of the confirmatory DBF gender DIF analyses for the studied item sets are presented in Table 4. As hypothesized, the results from the Form 1 analyses comparing matched male and female test takers confirm findings from previous research. For example, the beta value for the studied items in the Word Problem category is statistically significant indicating this item parcel is differentially easier for male test takers. The same pattern is present for other item categories: construction of Figures/ graphs, Figures/graphs present, Higher order thinking algebra, trigonometry, and analytic geometry items, as well as the geometry items. The item bundles are differentially more difficult for female testtakers. The beta value for the algebra operations item bundle is also statistically significant; this set of items favors females as hypothesized.

For the Form 2 analyses investigating comparable males and females testtakers, the pattern of results are similar to Form 1 results for several item categories: Word Problems, Figures/graphs present, Algebra Operations. The results from the Form 2 analyses for the rest of the studied item sets tended to be different. In contrast to the results from Form 1, the results of the SIBTEST analyses for the Construction of figures/graphs item parcels, and Higher Order Thinking items (algebra, trigonometry, and analytic geometry) and the geometry item parcels were not statistically significant. However, these categories do overlap. For example, items 38 and 39 is contained in both Construction of Figures/graphs and Higher Order Thinking analytic geometry items. Nevertheless, these studied item sets were not differentially easier for males on Form 2.

Insert Table 4

Exploratory DBF Analyses

Table 5 presents the results from the exploratory DIF analyses investigating whether male or female testtakers are differentially affected by variation in item position. The same item sets are the same as those listed in Table 5. Only the results from three analysis were statistically significant. The studied items in the Algebra operations category were differentially easier for the Form 2 females testtakers in comparison to the Form 1 female examinees. The algebra operations items were also differentially easier for the men completing Form 2. In addition, the analytic geometry item parcel was differentially more difficult for the women who took Form 1 of the test.

Insert Table 5

DISCUSSION

The purpose of this investigation was to examine whether changes in item position influences the patterns of gender DIF present in item types that tend to function differentially. Second, whether female or male testakers' performance is differentially affected by variation in item position was examined.

There were changes in the amount of gender DIF present when the confirmatory SIBTEST results for Form 1 male and female testtakers are compared to the results for Form 2 male and female examinees. The studied items for several categories, Construction of figures/graphs item parcels, and Higher Order thinking items (algebra, trigonometry, and analytic geometry) did not function differentially for the Form 2 men and women. The results from the exploratory DBF analyses suggest the analytic geometry items were differentially easier for women on Form 2.

M. Linn (1992) suggested that women may be at a time disadvantage taking standardized tests like the SAT-M since they tend not to take shortcuts in answering the test items. Gallagher (1992) did find females were more likely than males to use algorithms to solve math problems on the SAT-M. Using algorithms to solve problems can take more time than test-taking strategies like working backwards from the multiple choice options. While the exam studied in this investigation is not parallel to the SAT-M, the testing conditions, such as some multiple choice formats and time constraints are similar. Perhaps women saved time when answering items within content area on Form 2. Consequently, they had more time for answering the items at the analytic geometry items at the end of the test and test items in general.

Nevertheless, the findings from this investigation should be considered as preliminary and limited. First, the students taking this exam were admitted to a university with a highly-competitive admissions policy. The general college-bound population is likely to be more heterogeneous. Second, the number of items studied in this investigation and the available samples were limited. As a consequence, no cross validation study was conducted; the findings were not replicated within the study. While the sample sizes in the study should be adequate to obtain stable estimates (N=500) (Shealy & Stout, 1993), that may not be the case. Finally, determining the most appropriate matching criterion or "valid" subtest was not addressed in this study.

Nevertheless, more research in this area may be of interest. The notion of DBF is a useful addition for the study of differential item functioning. This approach provides added power for detecting patterns of differences. If DIF/DBF is conceptualized as differences between the reference and focal group on "nuisance" dimensions (See Ackerman, 1992; Stout, 1993), using DBF analyses to confirm the presence of "nuisance dimensions" is valuable.

Table 1

Descriptive Statistics for Advanced Mathematics

Form	Sample	N	Mean	Std Dev	Min	Max
1	Total	3932	21.07	8.07	2.00	40.00
	Males	2019	23.46	8.14	2.00	40.00
	Females	1877	18.56	7.15	3.00	40.00
2	Total	1074	21.60	8.72	3.00	40.00
	Males	554	24.14	8.79	4.00	40.00
	Females	511	18.93	7.80	3.00	39.00

Table 2

Dimtest Results for Advanced Mathematics Exam: (Forms Combined)

AT1	AT2&PT	T	P-value
Geometry	Algebra	-0.27	0.61
Analytic	Algebra & Geometry	-0.73	0.77
Trig	Algebra, Geometry Analytic Geometry	5.07	0.00**

Note: Algebra items (1-18) Geometry items (19-23)
Trig items (24-35) Analytic Geometry items (36-40)

Table 3
Differential Bundle Functioning Analyses Design

Set	Item Category (Attribute)	Test Form	Studied Items	Reference Group (N)	Focal Group (N)	Hypothesis
A	Word Problems(1,5)	1	Items 1,16,21,28,31	Males (2019)	Females (1877)	Easier for Males
B		2	Items 1,4,5,17,33	Males (554)	Females (511)	Easier for Males
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———
A	Figures/Graphs Present (4,11)	1	Items 12,17,18,19,32,36	Males (2019)	Females (1877)	Easier for Males
B		2	Items 19,20,21,23,26,37	Males (554)	Females (511)	Easier for Males
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———
A	Construction of Graphs/Figures (4,12)	1	Items 10,20,22,23,30,33	Males (2019)	Females (1877)	Easier for Males
B		2	Items 18,22,34,36,38,39	Males (554)	Females (511)	Easier for Males
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———

Table 3 (con't.)

Set	Item Category (Attribute)	Test Form	Studied Items	Reference Group (N)	Focal Group (N)	Hypothesis
A	Algebra Operations (1,3,9)	1	Items 5,8,26,29,34,39	Males (2019)	Females (1877)	Easier for Females
B		2	Items 2,6,11,13,15,16	Males (554)	Females (511)	Easier for Females
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———
A	Higher Order Thinking Algebra (16)	1	Items 6,23,28	Males (2019)	Females (1877)	Easier for Males
B		2	Items 12,17,18	Males (554)	Females (511)	Easier for Males
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———
A	Higher Order Thinking Trigonometry (10,16)	1	Items 4,21,22	Males (2019)	Females (1877)	Easier for Males
B		2	Items 32,33,34	Males (554)	Females (511)	Easier for Males
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———

Table 3 (con't.)

Set	Item Category (Attribute)	Test Form	Studied Items	Reference Group (N)	Focal Group (N)	Hypothesis
A	Higher Order Thinking	1	Items 10,20,30,36	Males (2019)	Females (1877)	Easier for Males
B	Analytic Geometry (3,4,8,10)	2	Items 36,37,38,39	Males (554)	Females (511)	Easier for Males
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———
A	Geometry	1	Items 12,17,18,32,33	Males (2019)	Females (1877)	Easier for Males
B		2	Items 19,20,21,22,23	Males (554)	Females (511)	Easier for Males
C		1 and 2	See Form 1 and 2 Items Above	Form 1 Males (2019)	Form 2 Males (554)	———
D		1 and 2	See Form 1 and 2 Items Above	Form 1 Females (1877)	Form 2 Females (511)	———

23

24

Table 4
Confirmatory Differential Bundle Functioning Analyses: Sets A and B

Content	Item	Beta-uni	SIB-uni z-statistic	SIB-uni p-value
Word Problems (1,5)				
Form 1	(1,16,21,28,31)	.461	12.498	.000
Form 2	(1,4,5,17,33)	.385	5.356	.000
Construction of Figures/graphs (4,12)				
Form 1	(10,20,22,23,30,33)	.112	2.656	.004
Form 2	(18,22,34,36,38,39)	.025	.290	.386
Figures/graphs present (4,11)				
Form 1	(12,17,18,19,32,36)	.103	2.410	.008
Form 2	(19,20,21,23,26,37)	.195	2.351	.009
Algebra Operation (1,3,9)				
Form 1	(5,8,26,29,34,39)	-.182	-4.531	.000
Form 2	(2,6,11,13,15,16)	-.166	-2.177	.015
Higher Order Thinking Algebra Items (16)				
Form 1	(6,23,28)	.096	3.564	.000
Form 2	(12,17,18)	.061	1.184	.118
Higher Order Thinking Trig Items (10,16)				
Form 1	(4,21,22)	.141	4.968	.000
Form 2	(32,33,34)	.049	.907	.182
Higher Order Thinking Analytic Geometry Items (3,4,8,10)				
Form 1	(10,20,30,36)	.068	1.971	.024
Form 2	(36,37,38,39)	.017	.257	.398
Geometry Items				
Form 1	(12,17,18,32,33)	.110	2.811	.002
Form 2	(19,20,21,22,23)	.065	.842	.200

Table 5
Results for Exploratory Differential Bundling Functioning Analyses: Sets C and D

Content	Item	Beta-uni	SIB-uni z-statistic	SIB-uni p-value
Word problem (1,5)				
Form1 VS Form2(male)		.068	1.265	.206
Form1 VS Form2(female)		-.036	-.626	.532
Construction of figures/graphs (4,12)				
Form1 VS Form2(male)		.014	.219	.826
Form1 VS Form2(female)		-.101	-1.582	.114
Figures/graphs present (4,11)				
Form1 VS Form2(male)		-.023	-.386	.700
Form1 VS Form2(female)		.119	1.776	.076
Algebra Operation (1,3,9)				
Form1 VS Form2(male)		-.191	-3.420	.000**
Form1 VS Form2(female)		-.140	-2.205	.028*
Higher Order Thinking Algebra Items (16)				
Form1 VS Form2(male)		.003	.082	.934
Form1 VS Form2(female)		-.025	-.635	.526
Higher Order Thinking Trig Items (10,16)				
Form1 VS Form2(male)		.015	.345	.730
Form1 VS Form2(female)		-.062	-1.561	.118
Higher Order Thinking Analytic Geometry Items (3,4,8,10)				
Form1 VS Form2(male)		-.056	-1.184	.236
Form1 VS Form2(female)		-.103	-2.064	.038*
Geometry Items				
Form1 VS Form2(male)		.054	1.008	.314
Form1 VS Form2(female)		.055	.933	.352

Note. The studied items are the same as Table 4.

Bibliography

Bridgman, B., & Lewis, C. (1996). Gender differences in college mathematics grades and SAT-M scores: A reanalysis of Wainer and Steinberg, Journal of Educational Measurement, 33(3), p, 257-270.

Burton, N. (1995, April). How have the changes in the SAT affected women's mathematics performance? Educational Measurement: Issues and Practice, 15(4),p, 5-9.

Doolittle, A.E., & Cleary, T.A. (1987). Gender-based differential item performance in mathematics achievement items, Journal of Educational Measurement, 24, 157-166.

Douglas, J., Roussos, L., & Stout, W. (in press). Item bundling DIF hypothesis testing: Identifying suspect bundles and assessing their DIF, Journal of Educational Measurement, 33(4), p, 465-484.

Gallagher, Ann M. (1992). Sex differences in problem-solving strategies used by high-scoring examinees on the SAT-M. (Research Report No 92-2) Princeton: Educational Testing Service.

Hambleton, R.K., & Traub, R.E. (1974). The effects of item order on test performance and stress. Journal of Experimental Education, 43, 40-46.

Harnisch, D.L., Tatsuoka, K.K., & Wilkins, J. (1995, November). Reporting math proficiencies based on the new SAT-M, paper presented at the American Evaluation Meeting, Vancouver, British Columbia, CANADA

Harris, A.M., & Carlton, S.T. (1993). Patterns of gender differences on mathematics items on the Scholastic Aptitude Test, Applied Measurement in Education, 6, 137-152.

Holland, P. W. & Thayer, D. T. (1986). Differential item performance and the Mantel-Haenszel procedure. (Research Report No 86-31) Princeton: Educational Testing Service.

Lane, S., Wang, N., & Magone, M. (1995, April). Gender-related differential item functioning on a middle-school mathematics performance assessment, Educational Measurement: Issues and Practice, 15(4), p, 21-27

Linn, M., (1992, October). Gender differences in educational achievement, Paper presented at the Fifty-second ETS Invitation Conference: Sex equity in educational opportunity, achievement, and testing, Princeton, NJ.

Nandakumur, R. (1993). Simultaneous DIF amplification and cancellation: Shealy-Stout's test for DIF, Journal of Educational Measurement, 30, 293-312.

Nandakumur, R. (1991). Traditional versus essential unidimensionality, Journal of Educational Measurement, 28, 99-118.

O'Neill, K.A., & McPeek, W.M. (1993). Item and test characteristics that are associated with differential item functioning. In P.W. Holland & H. Wainer (Eds.), Differential item functioning. Hillsdale, NJ: Lawrence Erlbaum Associates.

Plake, B.S., Ansorge, C.J., Parker, C.S., & Lowry, S.R. (1982). Effects of item arrangement, knowledge of arrangement, test anxiety, and sex on test performance, Journal of Educational Measurement, 19, 49-58.

Plake, B.S., Patience, W.M., & Whitney, D.R., (1988). Differential item performance in mathematics achievement test items: Effect of item arrangement, Educational and Psychological Measurement, 48, 885-894.

Ryan, K. & Fan, M. Validation studies for UIUC German, Spanish, and French Placement Examinations: Form 2", Measurement and Evaluation Division, Office of Instructional Resources, University of Illinois, Urbana, IL, Technical report prepared for American College Testing (ACT), Iowa City, IA under contract with the University of Illinois, (July, 1993).

Ryan, K.E. & Fan, M., (1996). Examining gender DIF on a multiple-choice test of mathematics: A confirmatory approach, Educational Measurement: Issues and Practice, 15(4), p, 21-27.

Shealy, R. & Stout, W. (1993). An item response theory model for test bias and differential test functioning. In P. Holland and H. Wainer (eds.), Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. Psychometrika, 52, 589-617.

Stout, W. (1993, December). Item and test bias: Multidimensional IRT viewpoint and detection via SIBTEST. Paper presented at a colloquium in the Educational Psychology Department at University of Illinois, Urbana, IL.

Stout, W. & Roussos, L. (1995). SIBTEST User Manual, University of Illinois, Urbana, IL.

Tatsuoka, K.K. (1993). Item construction and psychometric models appropriate for constructed response models. In R. Bennett & W. Ward (Eds.), Construction versus Choice in Cognitive Measurement, 107-133. Hilldale, NJ: Erlbaum.

Wainer, H. & Steinberg, L.S. (1992). Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: A bidirectional validity study, Harvard Educational Review, 62, 323-326.

Wise, L.L., Chia, W.J., Park, R.K. (1989, March). Item position effects for test of work knowledge and arithmetic reasoning, Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Yen, W.M. (1980). The extent, causes, and importance of context effects on item parameters for two latent trait models, Journal of Educational Measurement, 297- 311.

Wick, R. (1991). Effects of item order and context on estimation of NEP. reading proficiency, Educational Measurement: Issues and Practice, 10(3), 10-16.

Appendix A
Revised Description of Attributes:

1 - Dealing with odd & even integers, prime numbers, factors, rational numbers, ordering, ratios, percentages, place value, powers, roots, and averages.

2 - Dealing with variable (addition and subtraction only), linear equations, linear algebraic expressions, signed-numbers, absolute values, irrational numbers.

3 - Dealing with higher-degree algebraic expressions, functions, sets, simple probability, combinatorics, modes and medians, exponents with variables.

4 - Dealing with perimeter, area and volume for triangles, circles, rectangles and other geometric objects. In analytic geometry, dealing with points, lines, in relation to a coordinate system.

5 - Translating word problems into arithmetic and algebraic expression(s). Can identify implicit variables and relationships. Dealing with real-world problems and real-world experiences.

6 - Restructuring problems into solvable forms. Choosing better, simpler or quicker strategies to solve problems. Choosing from rules, properties and theorems the better, simpler or quickest one to use.

7 - Recalling and interpreting knowledge based on definitions, properties or relationships from arithmetic, algebra, and geometry. Can perform computations in arithmetic, geometry, signed numbers, absolute value, median and mode.

8 - Applying mathematical rules and properties to solve equations (simultaneous); derive, factor and compute algebraic expressions.

9 - Skill with calculator. Work on the complicated algebra operations.

10 - Application of higher mental processes to solve problems. Sorting problems into implicit component parts and restructuring them in order to make the problem solvable.

11 - Working with figures, tables, and graphs.

12 - Generate figures and/or tables for problem solving.

13 - Understand the properties of the right triangle.

Appendix A (con't.)

14 - Can take advantage of the form of the test items and other test-taking methods without solving the problem in the manner intended by the item writer. Can solve a task by working backwards from the multiple-choice options.

15 - Working with problems having several steps. These steps may be explicit or implicit. Can establish subgoals of the problem, order, prioritize and execute the subgoals in a step-by-step fashion.

16 - Can comprehend sentences with negation, "at least", comparison, "must be", "could be", and with the relations of increasing and decreasing.

17 - Keeping track of what a question is asking, paying attention to detail. Identify constraints. Follow verbally written instructions, read complex, long sentences.

18 - Straight forward translations of verbal expressions into mathematical expressions where variable terms(s), constant(s), and needed operation(s) are readily apparent.

19 - Apply the relationships between the functions of trigonometry and angles and the functions of trigonometry. Example :

$$\sin^2\theta + \cos^2\theta = 1 \quad \text{or} \quad \sin\theta = \cos\left(\frac{\pi}{2} - \theta\right)$$

20 - Utilize the graphs to express the function of the trigonomy for the problem solving.

Adapted from Harnisch, D., Tatsuoka, K.K., & Wilkins, J.L. (1995, November). Reporting math proficiencies based on new SAT-M items, Paper presented at the 1995 American Evaluation Association Meeting, Vancouver, BC

Description of Mathematical Challenges:

1 - Dealing with odd & even integers, prime numbers, factors, rational numbers, ordering, ratios, percentages, place value, powers, roots, and averages.

2 - Dealing with variable (addition and subtraction only), linear equations, linear algebraic expressions, signed-numbers, absolute values, irrational numbers.

3 - Dealing with higher-degree algebraic expressions, functions,sets, simple probability, combinatorics, modes and medians, exponents with variables.

4 - Dealing with perimeter, area and volume for triangles, circles, rectangles and other geometric objects. In analytic geometry, dealing with points, lines, in relation to a coordinate system.

5 - Translating word problems into arithmetic and algebraic expression(s). Can identify implicit variables and relationships. Dealing with real-world problems and real-world experiences.

6 - Restructuring problems into solvable forms. Choosing better, simpler or quicker strategies to solve problems. Choosing from rules, properties and theorems the better, simpler or quickest one to use.

7 - Recalling and interpreting knowledge based on definitions, properties or relationships from arithmetic, algebra, and geometry. Can perform computations in arithmetic, geometry, signed numbers, absolute value, median and mode.

8 - Applying mathematical rules and properties to solve equations (simultaneous); derive, factor and compute algebraic expressions.

9 - Reasoning and logical thinking. Reasoning deductively from cause to effect. Spatial reasoning skills. Identify and understand necessary and sufficient conditions and apply them.

10 - Application of higher mental processes to solve problems. Sorting problems into implicit component parts and restructuring them in order to make the problem solvable.

11 - Working with figures, tables, and graphs. Can generate figures to facilitate problem-solving activities.

12 - Can take advantage of the form of the test items and other test-taking methods without solving the problem in the manner intended by the item writer. Can solve a task by working backwards from the multiple-choice options.

13 - Working with problems having several steps. These steps may be explicit or implicit. Can establish subgoals of the problem, order, prioritize and execute the subgoals in a step-by-step fashion.

14 - Can comprehend sentences with negation, "at least", comparison, "must be", "could be", and with the relations of increasing and decreasing.

15 - Answering questions formatted as grid-ins. Deriving solutions by a top down approach.

16 - Keeping track of what a question is asking, paying attention to detail. Identify constraints. Follow verbally written instructions, read complex, long sentences.

17 - Straight forward translations of verbal expressions into mathematical expressions where variable term(s), constant(s), and needed operation(s) are readily apparent.

From Harnisch, D., Tatsuoka, K.K., & Wilkins, J.L. (1995, November). Reporting math proficiences based on new AT-M items, Paper presented at the 1995 American Evaluation Association Meeting, Vancouver, BC

TM027635

ERIC

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An Examination of Item Context Effects, DIF, And Gender DIF

Author(s): Katherine E. Ryan & Shuwan Chiu

Corporate Source: University of Illinois

Publication Date: March, 1997

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

[X] Check here
For Level 1 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) *and* paper copy.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all Level 2 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] Check here
For Level 2 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature: Katherine E Ryan

Printed Name/Position/Title: Katherine Ryan, Head

Organization/Address: University of Illinois

Telephone: 2173333490

FAX: 2172444431

E-Mail Address: K-Ryan6@uiuc.edu

Date: 7/17/97

(over)

CUA

THE CATHOLIC UNIVERSITY OF AMERICA

Department of Education, O'Boyle Hall
Washington, D.C. 20064
800 464-3742 (Go4-ERIC)

June 1997

Dear ERIC Contributor:

Thank you for contributing materials about assessment, evaluation, research methods, or learning theory to the ERIC System. Your contribution helps make the ERIC database one of the most popular and useful products in education. As a token of our appreciation, please accept this packet of recent ERIC Digests. ERIC digests are short reports designed to help members of the educational community keep up-to-date with trends and new developments. While they are most often prepared for practitioners, digests can also target other audiences, including researchers, parents, and students. Digests are in the public domain and we encourage you to copy and redistribute them. Digests are also available at our web-site (http://ericae2.educ.cua.edu).

I would like to call your attention to recent key developments at our web-site. As a joint project with Texas A&M, we have posted a wonderful series of "How-to" papers. These are booklets on a range of measurement and statistical topics. We have also mounted the ERIC database along with a Search Wizard to help you formulate quality searches. The K12ASSESS-L listserv now has over 1,300 subscribers. Our pathfinder, *Assessment and Evaluation on the Internet,* has received a prestigious 5-star award from the Argus Clearinghouse for its coverage of what is on the internet. This summer we will be creating an on-line library of full-text documents (including newspaper articles, posted essays, and books) from across the internet. In addition we are starting an on-line journal on educational assessment. The big news for the ERIC System is that, starting late summer, you will be able to order and receive documents though the internet (see http://edrs.com/Press/PressReleases/P022197.htm).

On the back of this letter is a copy of the ERIC Document Reproduction Release Form. Please take a moment and send us any quality documents that are not in the system. We feel you have a professional responsibility to share your good work.

Sincerely,

Lawrence M. Rudner,
Director

Please send the ERIC Document Reproduction Release Form to:
The Catholic University of America
ERIC Clearinghouse on Assessment and Evaluation
O'Boyle Hall, Room 210
Washington, DC 20064

ERIC® Clearinghouse on Assessment and Evaluation